*M2 Research Internship*

# Leveraging Vision–Language Models for multi-modal information extraction for Knowledge Graph construction

## Context

The restoration or naturalization of hydro-ecosystems is a major challenge for the coming years in order to protect and preserve the quality and quantity of river water. The Rhine River, along its Franco-German borderline and more broadly, has undergone extensive human interventions over centuries: dams, embankments, industrialization, pollution, re-naturalization efforts, etc [2]. Many restoration works - both recent and historical - have generated large amounts of textual documentation (reports, archival documents, project plans, regulations, scientific articles) and visual material (maps, drawings, aerial/satellite imagery, photographs, cross-sectional charts). However, that material is often unstructured, scattered across institutions, in multiple languages, and not organized to support comparative analysis, learning, or decision-making effectively.

Meanwhile, AI capabilities, especially in vision-language modeling (VLM), which combines understanding of images and associated text, have matured significantly. There is an opportunity to apply VLMs [1] in the context of restoration to extract, organize, and structure knowledge in interpretable forms such as knowledge graphs [4], so that restoration experience (successful / less successful) can inform future restoration design, policy, monitoring, etc.

The German-French project TETRA[1] focuses on the development of an AI based toolbox and methodology for the water domain. Four partners are involved, two German and two French partners, two private companies (SEBA, Thalès), and two public research centers (Fraunhofer IOSB, ICube). The project is organized into several work packages. This internship will contribute to that AI based toolbox, by focusing on the problem of restoration knowledge extraction from multi-modal sources (text + images) and structuring that knowledge in machine-interpretable formats (e.g. a case base, knowledge graph, structured database).

## Scientific Objectives

This internship builds upon prior research on the extraction and structuring of restoration knowledge, in particular the study presented in [3] which focused on textual reports documenting restoration works on the Rhine. This proposal main objective is to extend this foundation to the visual domain, aiming to integrate images and texts within a unified multi-modal framework. The overall goal is to enhance the extraction, organization and interpretation of knowledge from heterogeneous restoration documentation through the use of Vision-Language Models (VLMs).

A first step is the identification and characterization of relevant documentary sources. These include historical and recent reports, project plans, maps, technical drawings, and images related to restoration interventions on the Rhine river. Particular attention will focus on the nature and diversity of the available imagery such as aerial or satellite images, annotated diagrams and to their associated textual components, including captions, descriptive notes, and analytical narratives. This preparatory phase will establish the corpus and metadata necessary for subsequent model adaptation.

The second and main research task involves applying and refining VLMs to extract complementary information from visual and textual data [1, 6, 5]. The VLMs should recognize and describe restoration structures, spatial configurations, and temporal stages (before, during, and after restoration) from images. They should extract objectives, methodologies, outcomes, and environmental parameters from text. A key scientific challenge lies in the multi-modal alignment of information linking visual elements and textual references to produce consistent and interpretable outcomes.

Building on these results, the internship will contribute to the enrichment of an already existing structured knowledge model (ontology) [3], describing restoration cases through key properties including intervention type, environmental context, methods, results, constraints, and costs. In addition to enriching the knowledge model, another key point is populating the knowledge model by constructing knowledge graphs with information extracted from images and text, ensuring querying, comparison, and visualization by researchers and practitioners.

Evaluation will address both technical and functional dimensions. Quantitative assessment will measure the precision and completeness of the multi-modal extraction, while qualitative evaluation will focus on the usefulness of the structured knowledge for supporting analysis and decision-making (e.g. identifying effective restoration strategies under specific conditions).

Finally, the enriched knowledge model and the models developed during the internship will be integrated into the TETRA toolbox.

[1] `https://anr.fr/Project-ANR-22-FAI2-0006`

The suggested work plan is as follows:

1. Review of the literature. Survey of sources (text + images) for Rhine restoration. Familiarization with relevant VLMs and tools.
2. Enrich the existing schema (ontology) for restoration cases. Prototype image + text extraction on a small sample.
3. Adapt VLM for domain: e.g. prompt design, fine-tuning, or instruction fine-tuning if needed, integrate alignment between image and text.
4. Construct Knowledge Graphs. Populate the case base.
5. Evaluation: extraction metrics. Prototype queries. Refine schema (ontology) or extraction as needed.
6. Writing the internship report.

## Work Environment

This internship will be carried out mainly in the SDC team of the ICube laboratory (CNRS UMR 7357) in the framework of the TETRA project. The internship will begin in spring 2026 for an estimated period of 4-6 months. The candidate will be welcomed into the ICube laboratory and supervised by Florence Le Ber and Franco Giustozzi.

## Candidate

**Specific knowledge:** Knowledge on data science methods, knowledge representation and reasoning, knowledge graphs. Languages: Python, java, owl/sparql. Interest in the application domain, ability to work with experts who are not computer scientists would be appreciated.
**Education:** Student about to graduate a Master or Engineer (Bac + 5) with a specialization in Computer Science.

## How to apply

The interested candidates must send an email to Florence Le Ber (`florence.leber@engees.unistra.fr`) and Franco Giustozzi (`franco.giustozzi@insa-strasbourg.fr`) with the following documents:

- A CV,
- A cover letter (max. 1 page), including the applicant's motivation for applying and a brief explanation of their academic background,
- A transcript of the available grades for the current year and the past year.

# References


[1] Florian Bordes and et.al. An introduction to vision-language modeling. *CoRR*, abs/2405.17247, 2024.

[2] Valentin Chardon, Cassandra Euzen, and Laurent Schmitt. Functional river restoration as a lever for adapting to climate change from an interdisciplinary emblematic showcase on the upper rhine. *Journal of Environmental Management*, 393:127151, 2025.

[3] Fethi Ghazouani, Franco Giustozzi, and Florence Le Ber. Llm-driven case-base populating for structuring and integrating restoration experiences. In Isabelle Bichindaritz and Beatriz López, editors, *Case-Based Reasoning Research and Development*, pages 67–80, Cham, 2025. Springer Nature Switzerland.

[4] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

[5] Chenyang Liu, Jiafan Zhang, Keyan Chen, Man Wang, Zhengxia Zou, and Zhenwei Shi. Remote sensing spatiotemporal vision–language models: A comprehensive survey. *IEEE Geoscience and Remote Sensing Magazine*, pages 2–42, 2025.

[6] Zilun Zhang, Haozhan Shen, Tiancheng Zhao, Bin Chen, Zian Guan, Yuhao Wang, Xu Jia, Yuxiang Cai, Yongheng Shang, and Jianwei Yin. Georsmllm: A multimodal large language model for vision-language tasks in geoscience and remote sensing. *CoRR*, abs/2503.12490, 2025.